Submitted to the Annals of Applied Statistics

FUNCTIONAL ENSEMBLE SURVIVAL TREE: DYNAMIC PREDICTION OF ALZHEIMER'S DISEASE PROGRESSION ACCOMMODATING MULTIPLE TIME-VARYING COVARIATES

By Shu Jiang* Yijun Xie† and Graham A. Colditz*

*Washington University in St. Louis** *and University of Waterloo*†

With the exponential growth in data collection, multiple time-varying biomarkers are commonly encountered in clinical studies, along with rich set of baseline covariates. This paper is motivated by addressing a critical issue in the field of Alzheimer's disease (AD) in which we aim to predict the time for AD conversion in people with mild cognitive impairment to inform prevention and early treatment decisions. Conventional joint models of biomarker trajectory with time-to-event data rely heavily on model assumptions and may not be applicable when the number of covariates is large. This thus motivated us to consider a functional ensemble survival tree framework to characterize the joint effects of both functional and baseline covariates in predicting disease progression. The proposed framework incorporates multivariate functional principal component analysis to characterize the changing patterns of multiple time-varying neurocognitive biomarker trajectories and then nest these features within an ensemble survival tree in predicting the progression of AD. We provide a fast implementation of the algorithm that accommodates personalized dynamic prediction that can be updated as new observations are gathered to reflect the patient's latest prognosis. The algorithm is empirically shown to perform well in simulation studies and is illustrated through the analysis of data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We provide implementation of our proposed method in `R` package `funest`.

1. Introduction. Alzheimer's disease (AD) is one of the most prevalent disease worldwide which leads to memory loss and dementia [Mattson, 2004, LaFerla et al., 2007, Rabin et al., 2019]. Early detection is critical due to the lack of disease-modifying agents for patients diagnosed with AD. Mild cognitive impairment (MCI) is defined as the transition stage between the clinically normal and dementia state where it involves memory and language loss that is considered greater than expected age-related changes [Mattson, 2004]. As a result, MCI patients are typically enrolled as the target population for early prognosis and evaluation of therapies trials [Ewers et al.,

Keywords and phrases: Dynamic prediction, Ensemble survival tree, Functional data analysis, Time-varying covariates

2012]. There is considerable interest in identifying biomarkers or combination of covariates, so that the likelihood of predicting the neurodegenerative pathology due to Alzheimer's disease for patients diagnosed with MCI can be greater. See Park et al. [2012], Ewers et al. [2012], Gomar et al. [2014] for example. Accurate and robust prediction of disease progression to AD is thus important and critical to move the field forward [Risacher et al., 2009].

Tremendous amounts of data are being collected in the hopes of finding significant factors that may be associated with AD progression. In the dataset that motivated this work, the Alzheimer's Disease Neuroimaging Initiative (ADNI), the focus was on the collection of longitudinal assessments, magnetic resonance imaging and positron emission tomography imaging measures, as well as other biomarkers from blood and cerebrospinal fluid [Cuingnet et al., 2011]. Of those covariates collected in the cohort, many are time-varying. For example, the cognitive change in preclinical AD is a series of cognitive tests which are measured at each patient visits. The potential for discovery would be much greater in incorporating all available patient-specific covariates in predicting the progression of AD. However, the challenges may arise from i) high dimensionality of the baseline covariates; ii) presence of multivariate time-varying biomarkers; iii) non-linear and complex relationship between the covariates and the time-to-event outcome. A natural question then is how to best utilize this information to improve prediction performance to inform prevention and early treatment decisions.

Existing methods in the literature, such as the joint model, hinges on the pre-specified model assumptions for both the time-varying biomarker and the survival outcome [Rizopoulos, 2012]. However the nature of time-varying biomarkers may vary under different clinical settings, making it difficult to identify a suitable model. For illustration purposes, we present in Figure 1, the raw longitudinal trajectories for two of the longitudinal cognitive measures for 50 randomly selected MCI patients in ADNI. We can see that both the Mini Mental State Examination (MMSE, left) and Functional Activities Questionnaire (FAQ, right) trajectories have changing patterns over time and are highly variable within and between patients. In addressing this concern, nonparametric methods such as splines or kernel smoothing, have been adopted in the literature for prediction using the denoised smoothed values of the biomarker trajectories [Wu and Chiang, 2000, Welsh et al., 2002]. More recently, functional approaches such as functional principal component analysis (FPCA), has become a popular alternative for modeling time-varying predictors due to its ability to use extracted features (changing patterns) in addition to the denoised smoothed values which will likely improve prediction [Ramsay and Silverman, 2004, Wang et al., 2016].

Examples of functional data analysis applied to time-to-event data include Yan et al. [2017, 2018], Kong et al. [2018]. However all of the aforementioned methods focuses on the dynamic prediction of time-to-event outcome with a single time-varying biomarker. As a result, Li and Luo [2019] recently proposed the use of multiple longitudinal biomarkers in predicting the disease progression. However, their method contingent on the proportional hazards model which may not be realistic and viable especially when the number of covariates is large.

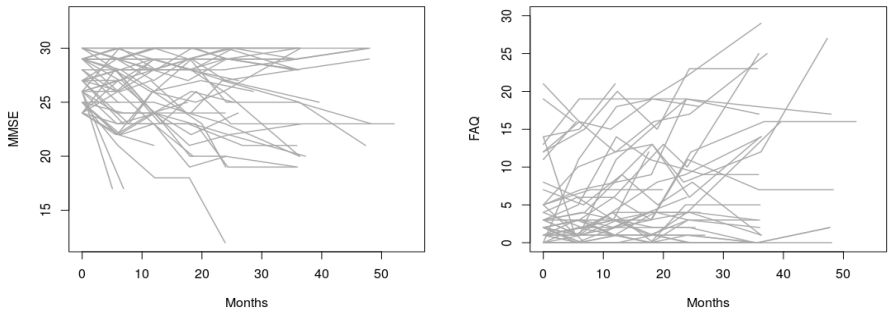

FIG 1. *Longitudinal trajectories of Mini Mental State Examination (MMSE, left) and Functional Activities Questionnaire (FAQ, right) of 50 randomly selected MCI patients in ADNI.*

In this article, we propose a unified strategy for dynamic prediction that does not depend on the model specification and can handle high dimensional baseline and multivariate time-varying covariates in the presence of right censoring. The proposed approach is entirely data-driven and can be stated in terms of three main steps. (1) First, extract features from the multivariate time-varying covariates such that the changing patterns can be summarized by a set of functional basis functions and the associated individualized functional scores. (2) Then, construct candidate estimators based on the extracted features and observed data. (3) Last, apply cross-validation to select the optimal estimator among all candidates in step 2. Specifically, we adopt tree-based methods in this paper where the possible candidate estimators in step 2 are generated by repeated binary recursive partitions [Ishwaran et al., 2008, 2011]. Tree-based methods facilitate a comprehensive modeling scheme and are appealing for their ability to handle data with high-dimensional covariates, facilitate complex and nonlinear relationship between predictors and outcomes and relax the proportional hazard assumption [Taylor, 2011, Jiang, 2019]. Given the tree-based estimators in

step 2, the optimal estimator in step 3 can be selected via cross-validation by tuning the number of functional basis functions from step 1 and tree-based parameters which we discuss in detail in Section 2. The proposed method will reinforce model robustness and prediction accuracy and serve as a valuable tool for researchers in conducting future research.

The remainder of this paper is organized as follows. In Section 2, we define notation and describe the model setup. In particular, we give detailed discussion on the multivariate principal component analysis (MFPCA) for feature extraction from multiple time-varying covariates and the construction of the functional ensemble survival tree for conducting individualized dynamic prediction. We investigate the finite sample performance in intensive simulation studies in Section 3 and provide publicly available code in `R` package `funest`. An application involving ADNI is given in Section 4, and concluding remarks and topics for future research are given in Section 5.

2. Notation and Method.

2.1. *Functional Ensemble Survival Tree.* Random survival forest (RSF) is an ensemble tree method that has been widely adopted for the analysis of right-censored survival data. The goal of constructing the RSF is to train a model that learns from the available functional and baseline covariates in the cohort, such that the model can be used to make risk predictions for new patients conditioning on partially observed data. The focus of this subsection is on model construction and we elaborate on individualized dynamic prediction in Section 2.2.

Typical RSF can not take longitudinal covariates directly as inputs. To extend the survival tree on the basis of longitudinal covariates, we first characterize the changing patterns of the time-varying biomarkers via MFPCA. We start by setting up the functional framework for single time-varying biomarkers and then expand to the multivariate setting. We let $Y_i = (Y'_{i,1}, \ldots, Y'_{i,Q})'$ be the observed time-varying biomarkers for individual i, $i = 1, \ldots, n$. The qth time-varying biomarker is denoted by $Y_{i,q} = (Y_{i,q}(t_{i,r}), \ldots, Y_{i,q}(t_{i,R_i}))'$ where R_i reflects random and irregular individual-specific visits, $q = 1, \ldots, Q$. We assume that the qth observed trajectory, $\forall q \in \{1, \ldots, Q\}$, is recorded with error,

$$Y_{i,q}(t_{i,r}) = Z_{i,q}(t_{i,r}) + \epsilon_{i,q,r} \ , \quad \forall t_{i,r} \in [0, \tau] \tag{2.1}$$

where $Z_{i,q}(t_{i,r})$ denotes the denoised mean value of $Y_{i,q}(t_{i,r})$ for $t_{i,r} \in [0, \tau]$ and τ denotes the maximum follow up time in the cohort. The error term is assumed to have $E(\epsilon_{i,q,r}) = 0$ and $var(\epsilon_{i,q,r}) = \sigma_q^2$ where $t_{i,r}$, $Z_{i,q}$ and $\epsilon_{i,q,r}$ are assumed to be mutually independent [Yao et al., 2005].

Under the functional framework, we assume that $Z_{i,q} = \{Z_{i,q}(t), \forall t \in [0,\tau]\}$ are realizations of a stochastic process $Z_q(t)$ in a square integrable functional space with domain τ. The stochastic process is assumed to have mean function $E[Z_q(t)] = \mu_q(t)$ and covariance operator $C_q(t,s) = Cov(Z_q(t), Z_q(s))$ for $\forall t, s \in [0,\tau]$. Then by Mercer's theorem [Mercer, 1909],

$$C_q(t,s) = \sum_{j=1}^{\infty} \lambda_j^q \phi_j^q(s)\phi_j^q(t), \tag{2.2}$$

where $\phi_j^q(t)$ is the jth orthonormal eigenfunction and λ_j^q is the corresponding eigenvalue where $\lambda_1^q \geq \lambda_2^q, \geq \ldots > 0$, $j = 1, \ldots, \infty$. This decomposition thus allows us to characterize each functional observation $Z_{iq}(t)$ as

$$Z_{i,q}(t) = \mu_q(t) + \sum_{j=1}^{\infty} \xi_{i,j}^q \phi_j^q(t), \tag{2.3}$$

where $\xi_{i,j}^q = \langle Z_{i,q}(t) - \mu_q(t), \phi_j^q(t)\rangle = \int_{t=0}^{\tau}[Z_{i,q}(t) - \mu_q(t)]\phi_j^q(t)dt$, is the jth functional principal component (FPC) score for individual i. According to the Karhunen–Loève theorem [Ramsay and Silverman, 2004], each curve $Z_{i,q}(t), \forall t \in [0,\tau]$, can then be characterized by the infinite sequence of FPC scores $\xi_{i,j}^q$, $j = 1, \ldots, \infty$. In practice, an approximation of (2.3) is usually carried out by truncating the infinite summation to the first M_q terms where M_q could be determined by, for example, Akaike information criterion (AIC) or the total variance explained (TVE) [Wang et al., 2016]. For estimation, given the observed data, we adopt the Principal Analysis by Conditional Estimation (PACE) algorithm for its well-known property of accommodating sparse longitudinal observations as is the case in our motivating study [Yao et al., 2005]. Specifically, we use PACE algorithm to facilitate the estimation of the discretized mean function $\hat{\mu}_q^{(i)} = (\hat{\mu}_q(t_{i,1}), \ldots, \hat{\mu}_q(t_{i,R_i}))'$, the $R_i \times R_i$ empirical covariance matrix $\hat{\Sigma}_i^q$ and the corresponding eigenvectors $\hat{\phi}_{i,j}^q$ and eigenvalues $\hat{\lambda}_{i,j}^q$, $j = 1, \ldots, M_q$. Then the univariate FPC scores for the qth biomarker trajectory for ith individual can be estimated as

$$\hat{\xi}_{i,j}^q = \hat{\lambda}_{i,j}^q (\hat{\phi}_{i,j}^q)^T (\hat{\Sigma}_i^q)^{-1}(Y_{i,q} - \hat{\mu}_q^{(i)}) \,, \tag{2.4}$$

$j = 1, \ldots, M_q$, for $q = 1 \ldots, Q$.

Next we combine the Q univariate time-varying biomarkers via MFPCA following Happ and Greven [2018]. We let $M = \sum_{q=1}^{Q} M_q$ and $\hat{\Lambda} \in \mathbb{R}^{n \times M}$ be an $n \times M$ matrix for which the ith row is $\{\hat{\xi}_{i,1}^1, \ldots, \hat{\xi}_{i,M_1}^1, \ldots, \hat{\xi}_{i,1}^Q, \ldots, \hat{\xi}_{i,M_Q}^Q\}$. In the multivariate setting we aim to perform a matrix eigenanalysis such

that we can estimate the corresponding eigenvectors $\hat{v}_m$, from the empirical block matrix $\hat{G} = \frac{1}{n-1}\hat{\Lambda}^T\hat{\Lambda} \in \mathbb{R}^{M\times M}$, $m = 1, \ldots, M$. Note that MFPCA indirectly accommodates the potential correlations among multiple trajectories via correlation among the FPC scores by pooling all estimated eigenvectors from the univariate biomarkers in the block matrix $\hat{G}$. The eigenvectors $\hat{v}_m$ thus contain the information of correlations across different time-varying biomarkers. As a result, the multivariate eigenfunctions are estimated as

$$\hat{\psi}_m^q(t_q) = \sum_{k=1}^{M_q} [\hat{v}_m]_k^q \hat{\phi}_k^q(t_q), \ t_q \in \tau, \tag{2.5}$$

where $[\hat{v}_m]_k^q$ denotes the kth entry in the qth block of $\hat{v}_m$, $q = 1, \ldots, Q$, $m = 1, \ldots, M$. The corresponding individual-specific MFPC scores can thus be estimated as

$$\hat{\rho}_{i,m} = \sum_{q=1}^{Q}\sum_{k=1}^{M_q} [\hat{v}_m]_k^q \hat{\xi}_{i,k}^q, \tag{2.6}$$

$m = 1, \ldots, M$. Similar to the univariate setting, the optimal number of MFPCs, $\{D : D \leq M\}$, can be chosen based on, for example, TVE or AIC.

The RSF can be easily constructed once the MFPCA scores have been estimated as in (2.6). Within the forest, every tree in the forest is grown from a single node to a tree with multiple terminal nodes. Specifically, each decision tree is grown by partitioning individuals at each node into two groups, where the split is chosen under a user-specified splitting rule. Node splitting rules often are determined with the goal to either maximize within-node homogeneity or between-node heterogeneity. The standard split criterion for survival trees is the log-rank statistic to maximize the survival differences at each node which has been widely used and implemented [Ishwaran et al., 2011]. Other splitting criterion such as the maximally selected rank statistic has been recently developed for its well-known unbiased split variable selection property [Wright and Ziegler, 2015]. In each terminal node of a tree, the survival function is estimated using the Kaplan–Meier estimator, utilizing only the observations from the same terminal node. Note that several parameters needs to be tuned via cross-validation. In particular, the prediction error needs to be assessed with, for example, various number of trees, number of covariates to split on and the minimal terminal node size. In addition, we may also tune the number for MFPCs that are nested within the RSF for a better prediction performance. See Section 3 for more details.

2.2. *Individualized Dynamic Prediction.* We let n denote the number of individuals in the training cohort and $n+1$ be the new individual who is event-free and has observation up to some time $t^\star, t^\star < \tau$. For each single tree b, $b = 1, \ldots, B$, prediction of the survival probability at $t^\star + \triangle t < \tau$, is made by dropping the new individual $n+1$'s observations down the tree as

$$\hat{S}_b(t^\star + \triangle t | t^\star) = \frac{\hat{S}_b(t^\star + \triangle t | W_{n+1}, \hat{\rho}_{n+1})}{\hat{S}_b(t^\star | W_{n+1}, \hat{\rho}_{n+1})} , \tag{2.7}$$

where W_{n+1} is the baseline covariates for individual $n+1$ of dimension $P \times 1$. The MFPC scores $\hat{\rho}_{n+1}$ can be obtained by first estimating the univariate FPC scores from (2.4),

$$\hat{\xi}^q_{n+1,j} = \hat{\lambda}^q_{n+1,j} (\hat{\phi}^q_{n+1,j})^T (\hat{\Sigma}^q_{n+1})^{-1} (Y_{n+1,q} - \hat{\mu}_q) , \tag{2.8}$$

$j = 1, \ldots, M_q$, $q = 1, \ldots, Q$. We then pass these FPC scores to (2.6) to obtain the MFPCA scores $\hat{\rho}_{n+1} = (\hat{\rho}_{n+1,1}, \ldots, \hat{\rho}_{n+1,D})'$. The final prediction from the forest is estimated by averaging over B trees in as

$$\hat{S}(t^\star + \triangle t | t^\star) = \frac{1}{B} \sum_{b=1}^{B} \hat{S}_b(t^\star + \triangle t | t^\star) . \tag{2.9}$$

3. Simulation Study. We conduct intensive simulation studies to investigate the finite sample performance of our proposed method in this section. We aim to mimic the motivating application and simulate $n = 400$ individuals in each dataset with $nsim = 500$. The individual-specific visit times $\{t_{i,r}, r = 1, 2, \ldots, 7\}$ are generated from the Gaussian distribution centered at $0, 3, 6, 9, 12, 15$, and 18 with standard deviation of 0.1 except the initial baseline visit which is fixed at 0.

We assume that the time-varying biomarkers are recorded with error, $Y_{i,q}(t_{i,r}) = Z_{i,q}(t_{i,r}) + \epsilon_{i,r,q}$, where $\epsilon_{i,r,q} \sim N(0,1)$ and $q = 1, 2, 3$. We consider both the linear and non-linear longitudinal trajectories in a similar fashion as Li and Luo [2019]. Specifically in the linear setting, we simulate

$$Z_{i,q}(t_{i,r}) = \beta_{0q} + \beta_{tq} t_{i,r} + \beta_{1q} X_{i,q} + b_{i,q},$$

where $[\beta_{01}, \beta_{02}, \beta_{03}] = [1.5, 2, 0.5]$, $[\beta_{t1}, \beta_{t2}, \beta_{t3}] = [1.5, -1, 0.6]$, and $[\beta_{11}, \beta_{12}, \beta_{13}] = [2, -1, 1]$. We simulate $X_{i,q} \sim N(3,1)$ for $q = 1, 2, 3$ and the individual-specific random effects $[b_{i,1}, b_{i,2}, b_{i,3}] \sim MVN(\mathbf{0}, \Sigma)$ with

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \eta_{12}\sigma_1\sigma_2 & \eta_{13}\sigma_1\sigma_3 \\ & \sigma_2^2 & \eta_{23}\sigma_2\sigma_3 \\ & & \sigma_3^2 \end{bmatrix} ,$$

where $[\sigma_1^2, \sigma_2^2, \sigma_3^2] = [1, 1.5, 2]$, and $[\eta_{12}, \eta_{13}, \eta_{23}] = [-0.2, 0.1, -0.3]$.

The nonlinear trajectories for each individual i is assumed to follow a piecewise model

$$Z_{iq}(t_{ir}) = \beta_{0q} + \beta_{tq} \sum_{r=1}^{3} c_r s^{(+)}(t_{ir} - k_r) + \sum_{q=1}^{3} \beta_{1q} X_{iq} + b_{iq},$$

where $[c_1, c_2, c_3] = [1.2, 0.7, 0.5]$, $[k_1, k_2, k_3] = [0, 6, 13]$, and

$$s^{(+)}(t) = \begin{cases} t, & t \leq 0 \\ 0, & \text{otherwise} \end{cases}.$$

We assume a porportional hazards model in this simulation where the conditional hazard function follows

$$h_i(t) = h_0(t) \exp\left\{ \sum_{p=1}^{P} \gamma_p W_{i,p} + \sum_{q=1}^{3} \alpha_q Z_{i,q}(t) \right\}$$

where α_q is set to be $(0.1, -0.1, 0.2)$ for $q = 1, 2, 3$ respectively. We consider four different scenarios for the set of fixed covariates $W_i = (W_{i,1}, \ldots, W_{i,P})'$. In the first two scenarios we set $\rho = 0.2$ and 0.5 for $P = 20, 100$ respectively to represent strong autoregressive dependence where $W_i \sim \text{MVN}(0, \Sigma^{(W)})$ with the (k, l)th component of $\Sigma^{(W)}$ define as

$$\Sigma^{(W)}_{k,l} = \begin{cases} 1, & k = l \\ \rho^{|k-l|}, & i \neq j \end{cases}.$$

In the last two scenarios, we consider binary covariates with $P(W_{i,p} = 1) = 0.5$ similarly under $\rho = 0.2$ and 0.5 for $P = 20, 100$. We set the associated coefficients $\gamma_p = (-2.5, -0.5, -0.15, -0.15, -0.1)$ for $p = 1, \ldots 5$ so that high values of $\gamma_1, \ldots \gamma_5$ are associated with shorter times to the event, and $\gamma_p = 0$ for $p = 6, \ldots, P$. The elements of W_i with non-zero coefficients were chosen to give both weak and strong dependence within the set of important covariates accompanied with set of noise variables. With the above setups, we are then ready to simulate $\tilde{T}_i$ and C_i. As demonstrated in Austin [2012] the survival time T_i can be generated from the inverse of the cumulative hazard function $H_i^{-1}(u|\mathcal{D}_i; \theta)$ where $H(t)$ is the cumulative hazard function and $u \sim \text{unif}(0, 1)$. We have simulated under the independent censoring scheme, where the censoring time is set to follow a uniform distribution $unif(0, C_m)$ where C_m is set such that the % of being censored by the end of the study is 30%.

In the simulations that we conducted, 300 individuals were randomly chosen from each simulated dataset to train the model and the 100 used to evaluate the prediction performance. To avoid overfitting, we employed a 5-fold inner and outer cross-validation. Specifically for the inner cross-validation, an optimal ensemble survival tree model was built and selected based on the best prediction performance by tuning the parameters in each fold. For each fold in the outer cross-validation, the prediction accuray measure is recorded dynamically for each time window $(t^\star, t^\star + \triangle t]$ conditional on data observed up to $t^\star$,$t^\star = 6, 9$, forecasting $t^\star + \triangle t$ for $\triangle t = 3, 6$.

Table 1 illustrates simulation results from the nonlinear setting. Additional simulation results under the linear setting are provided in Table 2 within the Supplemental Material for interested readers. As shown in Table 1, the AUC [Li et al., 2015] outputted from the proposed method are in good agreement with the true AUC confirming a satisfactory model discrimination. Additionally, we see that the Brier scores [Schoop et al., 2008] are very close to zero which confirms good model calibration as well. From our results, we can see that when the signal-to-noise ratio (S:N) decreases from $5 : 15$ to $5 : 95$, the proposed model retains robust performance in both the AUC and Brier score.

The proposed method has been implemented in our `funest` package which utilizes the well developed `ranger` that wraps the implementation of random forest in C++. The computational speed of our package is outstanding, as it takes less than 30 seconds for growing and making dynamic predictions on the functional ensemble survival forest with $\sim$2500 trees on a desktop with i7-7700 CPU. In addition, the package naturally takes advantage of the multi-core processor when running in a larger scale computational environment which warrants a even more promising computational speed.

4. Alzheimer's Disease Neuroimaging Initiative. The data used in this section were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database[1]. We treat the conversion from MCI to AD as the time-to-event outcome and focus on 317 patients who have been diagnosed with MCI in ADNI-1. Out of those who were diagnosed with MCI, 141 of them progressed to AD before the end of the study. Patients were assessed at baseline, 6, 12, 18, 24, and 36 months in ADNI-1 with additional annual follow-ups in ADNI-2 resulting in an average follow-up period of 33.4 (sd = 14.1) months. The corresponding average number of visits recorded was of 6.3 (sd = 2.3). Table 3 in the Supplemental Material shows the list of variables that we consider in this section. We have focused on five time-

[1] `http://adni.loni.usc.edu/`

TABLE 1

Estimated $AUC(t^\star, t^\star + \triangle t)$ and Brier score$(t^\star, t^\star + \triangle t)$ under the nonlinear setting via functional ensemble survival tree; $n = 400$, $nsim = 500$, S:N = signal-to-noise ratio.

W_i	P	S:N	$t^\star$	$\triangle t$	True AUC	AUC	BS
Normal	20	5:15	6	3	0.892	0.847	0.134
				6	0.904	0.864	0.147
			9	3	0.877	0.815	0.151
				6	0.896	0.830	0.147
	100	5:95	6	3	0.898	0.855	0.137
				6	0.913	0.870	0.142
			9	3	0.881	0.813	0.164
				6	0.899	0.827	0.157
Binary	20	5:15	6	3	0.827	0.781	0.085
				6	0.848	0.816	0.143
			9	3	0.843	0.805	0.123
				6	0.868	0.838	0.149
	100	5:95	6	3	0.836	0.792	0.083
				6	0.867	0.834	0.127
			9	3	0.854	0.819	0.114
				6	0.883	0.852	0.133

varying neurocognitive markers as well as other baseline covariates that have been well studied in the Alzheimer's literature [Mattson, 2004, LaFerla et al., 2007, Gomar et al., 2014].

Figure 2 illustrates the dynamic prediction performances under different models. In particular, we considered model 1 which is the full model that includes all available covariates and model 2 which includes only baseline covariates (including baseline measures for the time-varying markers). To avoid overfitting, we employed a 5-fold inner and outer cross-validation in this analysis similar to our simulation study. An optimal ensemble survival tree model was built and selected based on the best prediction performance by tuning the parameters in each fold in the inner cross-validation. For each fold in the outer cross-validation, the prediction accuracy measure is recorded dynamically for each time window $(t^\star, t^\star + \triangle t]$ conditional on data observed up to $t^\star$,$t^\star = 6, 12, 18, 24$ (month), forecasting $t^\star + \triangle t$ for $\triangle t = 6$

month. It is apparent that model 1 achieves a better AUC$(t^\star, t^\star + \triangle t)$ and BS$(t^\star, t^\star + \triangle t)$ dynamically over all time points. This suggests that the inclusion of the changing pattern of time-varying covariates indeed facilitates a better model discrimination and calibration.

We further illustrate the variable importance ranking via the variable permutation importance measure as shown in Figure 3. A variable is identified as important if it exerts a positive effect on the prediction performance. A greater value of permutation measure on a variable implies that the variable is more important for the overall predictive accuracy; see Nembrini et al. [2018] for more details. As a result, we see from Figure 3 that the first principal component (PC1) stands out with significantly large permutation importance measure in relation to other variables. This finding suggests that the contribution of the changing patterns of time-varying covariates in predicting the progression of AD for those who are diagnosed as MCI is much greater relative to the fixed baseline covariates. Such finding is indeed in agreement with what we observe in Figure 2.

Lastly, we demonstrate individualized dynamic prediction where we randomly set aside a single patient from the cohort. The model was trained on the remainder of the cohort leaving this single patient out, such that we are able to visualize the predicted future biomarker trajectory and risk conditional on partial profile. Figure 4 shows two of the five neurocognitive markers (i.e., ADAS-COG13 and FAQ) that we use in the model for ease of visualization. The dashed line in Figure 4 represents the last time the biomarker has been recorded for the patient. From the first column of Figure 4, we can see that the predicted trajectories of the ADAS-COG13 and FAQ are in great harmony with the true values. Correspondingly, the predicted AD-free probability for the patient is shown as a function of time in the second column where splines have been adopted to provide a smooth curve. The predicted low risk should be consistent with what one would expect from the stability of neurocognitive marker measurements. A full set of neurocognitive markers and their associated predictions are provided in Figure 5 in Supplemental Material for interested readers.

5. Discussion. We have formulated the functional ensemble survival tree framework that facilitates individualized dynamic prediction for disease progression accommodating multiple time-varying biomarkers. The proposed framework is fully data-driven and therefore removes the burden of the need to impose model assumptions on both the time-varying trajectories and the survival distribution. Specifically, we adopt MFPCA to characterize the changing pattern of the multivariate time-varying biomarkers which ef-

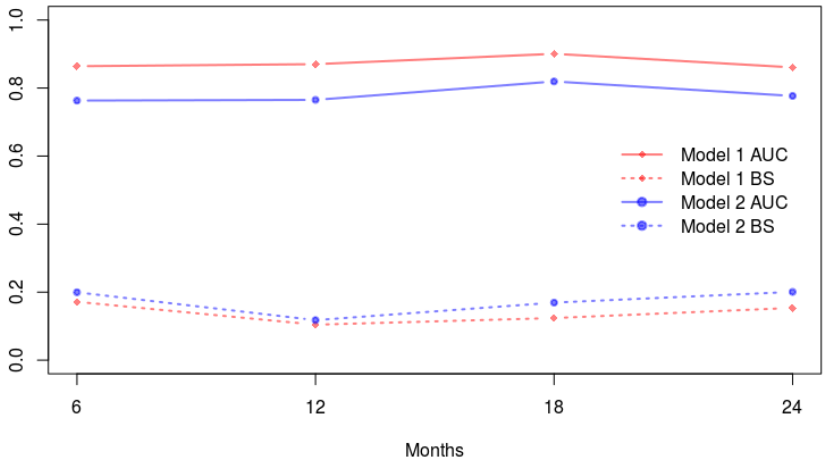

FIG 2. *Comparison of dynamic prediction performances of the full model (model 1) and baseline model (model 2) under a 5-fold cross-validation.* $AUC(t^\star, t^\star + 6)$ *and* $BS(t^\star, t^\star + 6)$ *conditions on data observed prior to* $t^\star = 6, 12, 18, 24$ *(month) in forecasting* $t^\star + 6$ *under a sliding window framework.*

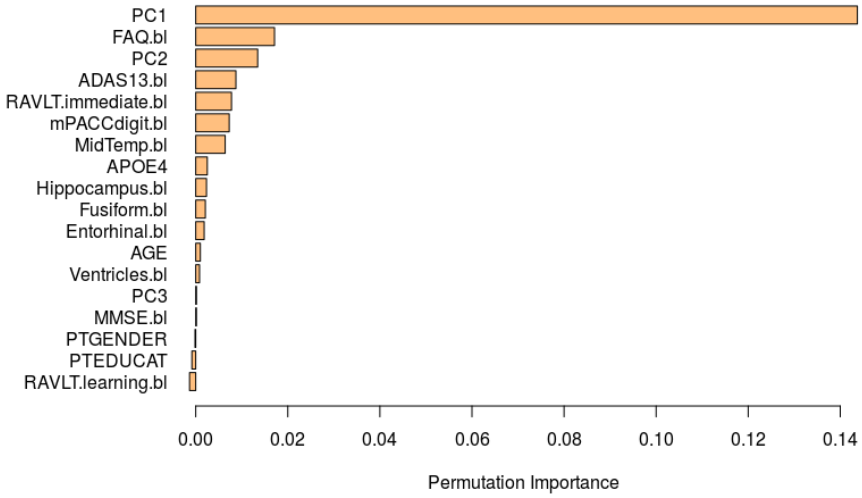

FIG 3. *Varbiable permutation importance barplot.*

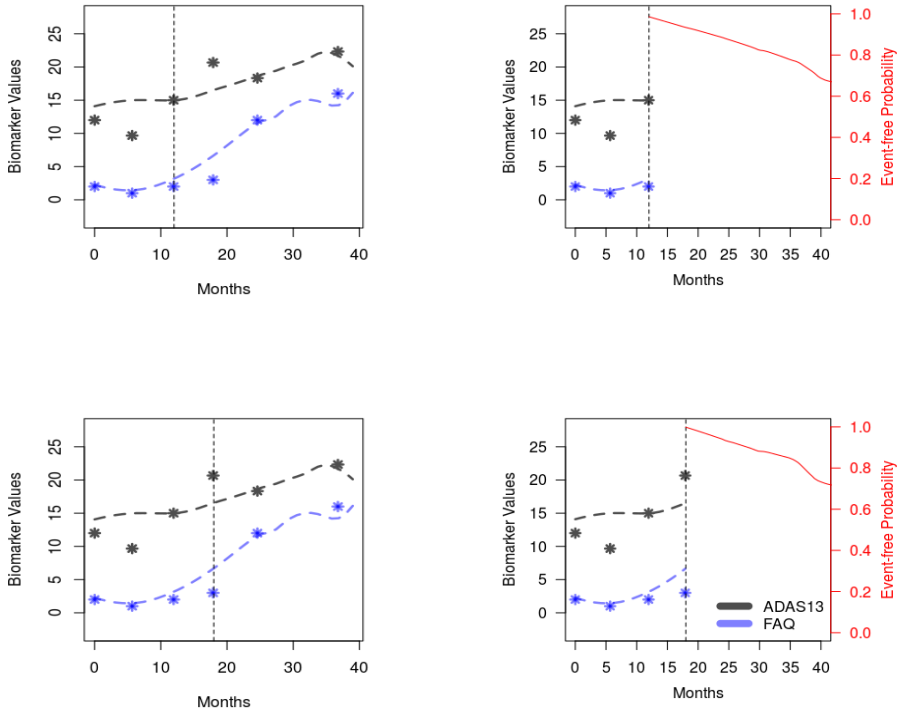

FIG 4. *Predicted trajectories of ADAS-COG13 and FAQ in the first column and predicted AD-free probability in the second column conditional on partially observed marker values prior to the dashed line; dashed line represents the last time the biomarker has been recorded for the patient.*

fectively captures the correlation among them. We also nest these extracted features into the ensemble survival tree which accommodates dynamic prediction under high dimensionality of baseline covariates and complex associations between the covariates and the time-to-event outcome. We investigate the empirical performance of the proposed algorithm and show that the model is robust and has a good discrimination and calibration via both the AUC and Brier score. We describe how to conduct individualized dynamic prediction and illustrate the proposed framework in the ADNI dataset. Furthermore, we make the proposed algorithm publicly available in `R` package `funest`. This could help physicians to predict future course of the biomarker trajectories as well as the associated risk of AD which in turn, could facilitate identifying high risk individuals for prevention trials and treatment

interventions.

A limitation in all tree-based methods is the lack of interpretability. However, in analyzing the ADNI dataset, we provided the variable permutation importance measure which identifies variable that are important contributors for the overall predictive accuracy. Our findings from ADNI suggest that time-varying trajectories play a major role in predicting AD progression for those that are diagnosed with MCI. We did not use the genetic marker data in our analysis as ADNI is an ongoing project that currently only has a sparse number of individuals who have their genetic profiles available. However, the model set up and the software distributed in this article warrants further research incorporating a richer set of covariates.

Acknowledgements. This work is supported in part by funding from the Foundation for Barnes Jewish Hospital and P30 CA091842. Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimag- ing Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/howtoapply/ADNIAcknowledgementList.pdf.

References.

Peter C Austin. Generating survival times to simulate cox proportional hazards models with time-varying covariates. *Statistics in medicine*, 31(29):3946–3958, 2012.

Rémi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, and Olivier Colliot. Automatic classification of patients with alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766 – 781, 2011. Multivariate Decoding and Brain Reading.

Michael Ewers, Cathal Walsh, John Q. Trojanowski, Leslie M. Shaw, Ronald C. Petersen, Clifford R. Jack, Howard H. Feldman, Arun L.W. Bokde, Gene E. Alexander, Philip Scheltens, Bruno Vellas, Bruno Dubois, Michael Weiner, and Harald Hampel. Prediction of conversion from mild cognitive impairment to alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging*, 33 (7):1203 – 1214.e2, 2012.

Jesus J. Gomar, Concepcion Conejero-Goldberg, Peter Davies, and Terry E. Goldberg. Extension and refinement of the predictive value of different classes of markers in ADNI: Four-year follow-up data. *Alzheimer's & Dementia*, 10(6):704 – 712, 2014.

Clara Happ and Sonja Greven. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659, 2018.

Hemant Ishwaran, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. Random survival forests. *Ann. Appl. Stat.*, 2(3):841–860, 2008.

Hemant Ishwaran, Udaya B. Kogalur, Xi Chen, and Andy J. Minn. Random survival forests for high-dimensional data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 4(1):115–132, 2011.

Shu Jiang. Prediction based on random survival forest. *American Journal of Biomedical Science & Research*, 6(2):109–111, 2019.

Dehan Kong, Joseph G. Ibrahim, Eunjee Lee, and Hongtu Zhu. Flcrm: Functional linear cox regression model. *Biometrics*, 74(1):109–117, 2018.

F LaFerla, K Green, and S Oddo. Intracellular amyloid- in alzheimer's disease. *Nat Rev Neurosci*, 8:499–509, 2007.

Kan Li and Sheng Luo. Dynamic prediction of alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data. *Statistics in Medicine*, 38(24):4804–4818, 2019.

Liang Li, Bo Hu, and Tom Greene. A simple method to estimate the time-dependent roc curve under right censoring. 2015.

M Mattson. Pathways towards and away from alzheimer's disease. *Nature*, 430:631–639, 2004.

James Mercer. Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, 209(441-458):415–446, 1909.

Stefano Nembrini, Inke R König, and Marvin N Wright. The revival of the gini importance? *Bioinformatics*, 34(21):3711–3718, 2018.

LQ Park, AL Gross, DG McLaren, and et al. Confirmatory factor analysis of the ADNI neuropsychological battery. *Brain Imaging and Behavior*, 6:528 – 539, 2012.

Jennifer S. Rabin, Hannah Klein, Dylan R. Kirn, Aaron P. Schultz, Hyun-Sik Yang, Olivia Hampton, Shu Jiang, Rachel F. Buckley, Anand Viswanathan, Trey Hedden, Jeremy Pruzin, Wai-Ying Wendy Yau, Edmarie Guzmán-Vélez, Yakeel T. Quiroz, Michael Properzi, Gad A. Marshall, Dorene M. Rentz, Keith A. Johnson, Reisa A. Sperling, and Jasmeer P. Chhatwal. Associations of Physical Activity and -Amyloid With Longitudinal Cognition and Neurodegeneration in Clinically Normal Older Adults. *JAMA Neurology*, 76(10):1203–1210, 2019.

JO Ramsay and BW Silverman. *Functional Data Analysis*. Springer Series in Statistics, 2004.

Shannon L Risacher, Andrew J Saykin, John D Wes, Li Shen, Hiram A Firpi, and Brenna C McDonald. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Current Alzheimer Research*, 6(4):347–361, 2009.

D Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data*. New York: Chapman and Hall/CRC, 2012.

R Schoop, E Graf, and M Schumacher. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics*, 64(2):603–610, 2008.

Jeremy M.G. Taylor. Random survival forests. *Journal of Thoracic Oncology*, 6(12):1974 – 1975, 2011.

Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295, 2016.

Alan H Welsh, Xihong Lin, and Raymond J Carroll. Marginal longitudinal nonparametric regression. *Journal of the American Statistical Association*, 97(458):482–493, 2002.

Marvin N Wright and Andreas Ziegler. ranger: A fast implementation of random forests for high dimensional data in c++ and r. *arXiv preprint arXiv:1508.04409*, 2015.

Colin O. Wu and Chin-Tsang Chiang. Kernel smoothing on varying coefficient models

with longitudinal dependent variable. *Statistica Sinica*, 10(2):433–456, 2000.

Fangrong Yan, Xiao Lin, and Xuelin Huang. Dynamic prediction of disease progression for leukemia patients by functional principal component analysis of longitudinal expression levels of an oncogene. *Ann. Appl. Stat.*, 11(3):1649–1670, 2017.

Fangrong Yan, Xiao Lin, Ruosha Li, and Xuelin Huang. Functional principal components analysis on moving time windows of longitudinal data: dynamic prediction of times to event. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(4): 961–978, 2018.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

SUPPLEMENTARY MATERIAL

Additional Simulation Results. The prediction performance measures under the linear setting in Section 3 are displayed in Table 2.

TABLE 2

Estimated $AUC(t^\star, t^\star + \triangle t)$ and Brier score$(t^\star, t^\star + \triangle t)$ under the linear setting via functional ensemble survival tree; $n = 400$, $nsim = 500$, S:N = signal-to-noise ratio.

W_i	P	S:N	$t^\star$	$\triangle t$	True AUC	AUC	BS
Normal	20	5:15	6	3	0.891	0.852	0.128
				6	0.909	0.869	0.146
			9	3	0.887	0.828	0.146
				6	0.910	0.846	0.136
	100	5:95	6	3	0.900	0.856	0.129
				6	0.918	0.876	0.140
			9	3	0.888	0.824	0.162
				6	0.908	0.839	0.155
Binary	20	5:15	6	3	0.751	0.678	0.066
				6	0.762	0.679	0.168
			9	3	0.760	0.689	0.141
				6	0.796	0.717	0.210
	100	5:95	6	3	0.759	0.663	0.060
				6	0.775	0.690	0.139
			9	3	0.762	0.679	0.127
				6	0.792	0.703	0.190

Additional Real Data Results. Table 3 gives a full list of the covariates that we have used in Section 4 for the ADNI dataset. Additional individualized dynamic prediction plots with all neurocognitive markers are displayed in Figure 5.

().

TABLE 3
Covariates used in ADNI dataset; † represents a time-varying covariate.

Variable	Description
ADAS-Cog 13†	Alzheimer Disease Assessment Scale-Cognitive 13 items
RAVLT.immediate†	Rey Auditory Verbal Learning Tests immediate score
RAVLT.learning†	Rey Auditory Verbal Learning Tests learning score
FAQ†	Functional Assessment Questionnaire
MMSE†	Mini Mental State Examination
mPACCdigit.bl	baseline preclinical Alzheimer's cognitive composite score
MidTemp.bl	baseline Middle temporal gyrus volume
Hippocampus.bl	baseline Hippocampus gyrus volume
Fusiform.bl	baseline Fusiform gyrus volume
Entorhinal.bl	baseline Entorhinal volume
Ventricles.bl	baseline Ventricles volume
ADAS-Cog13.bl	baseline Alzheimer Disease Assessment Scale-Cognitive 13 items
RAVLT.immediate.bl	baseline Rey Auditory Verbal Learning Tests immediate score
RAVLT.learning.bl	baseline Rey Auditory Verbal Learning Tests learning score
FAQ.bl	baseline Functional Assessment Questionnaire
MMSE.bl	baseline Mini Mental State Examination
APOE4	apolipoprotein E gene indicator
AGE	age at recruitment
PTGENDER	gender
PTEDUCAT	years of education

SHU JIANG
DIVISION OF PUBLIC HEALTH SCIENCES
DEPARTMENT OF SURGERY
WASHINGTON UNIVERSITY IN ST. LOUIS
ST. LOUIS, MO, USA
E-MAIL: jiang.shu@wustl.edu

YIJUN XIE
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCES
UNIVERSITY OF WATERLOO
WATERLOO, ON, CANADA
E-MAIL: yijun.xie@uwaterloo.ca

GRAHAM A. COLDITZ
DIVISION OF PUBLIC HEALTH SCIENCES
DEPARTMENT OF SURGERY
WASHINGTON UNIVERSITY IN ST. LOUIS
ST. LOUIS, MO, USA
E-MAIL: colditzg@wustl.edu

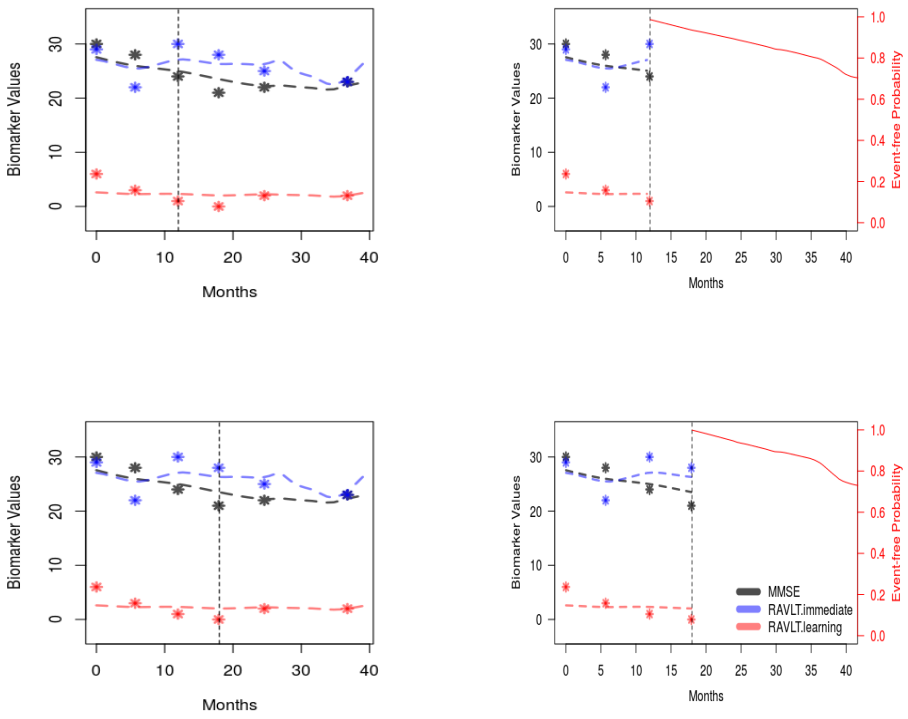

FIG 5. *Predicted trajectories of MMSE, RAVLT.immediate and RAVLT.learning in the first column and predicted AD-free probability in the second column conditional on partially observed marker values prior to the dashed line; dashed line represents the last time the biomarker has been recorded for the patient.*